Amendments to the Claims

Please cancel Claims 1, 4, 5, 10, 12-14, 19-21, 26, 35 and 36. Please amend Claims 2, 3, 7, 8, 11, 16, 22, 24, 27, 29 and 34. Please add new Claim 37. The Claim Listing below will replace all prior versions of the claims in the application:

Claim Listing

1. (Cancelled)

2. (Currently amended) A method as claimed in Claim [[1]] 34 wherein the set of known biological fragments is from published databases of motifs or proteins.

3. (Currently amended) A method as claimed in Claim [[1]] 34 ~~further comprising the step of:~~ wherein step (f) includes providing ~~for each desired~~ a plurality of subject genome sequences, ~~using said set of known biological fragments, repeating the comparing and forming steps such that a respective vector representation is formed~~ and step (h) forms a respective feature vector for each subject genome sequence such that each ~~desired~~ subject genome sequence has a respective vector representation of a same length, said set of known biological fragments being a same set used for all of said subject genome sequences.

4 - 6. (Cancelled)

7. (Currently amended) A method as claimed in Claim [[1]] 34 wherein the subject genome sequence is a DNA sequence or subsequence or protein sequence or subsequence.

8. (Currently amended) A method as claimed in Claim [[1]] 34 wherein step (g) quantitatively determining a score ~~the counting~~ includes determining probability of the subject genome sequence being generated by the known biological fragment by (i) counting the number of times the known biological fragment is found in the subject genome sequence and (ii) from said counted number of times, forming a vector element,

such that for each known biological fragment there is a respective vector element representing the number of times that known biological fragment is found in the subject genome sequence.

9. (Original) A method as claimed in Claim 8 wherein the counting determining probability employs a 0-th order Markov model for each known biological fragment.

10. (Cancelled)

11. (Currently amended) Apparatus as claimed in Claim [[10]] 37 wherein the data store is a published database of motifs or proteins.

12 - 15. (Cancelled)

16. (Currently amended) Apparatus as claimed in Claim [[10]] 37 wherein the subject genome sequence is a DNA sequence or subsequence or protein sequence or subsequence.

17 - 21. (Cancelled)

22. (Currently amended) The method of Claim [[21]] 34 wherein the respective representation of each known biological fragment is a text string.

23. (Previously Presented) The method of Claim 22 wherein quantitatively determining a score of each known biological fragment in the set includes for each known biological fragment, counting the number of times the text string of the respective representation is found within the subject genome sequence.

24. (Currently amended) The method of Claim [[21]] 34 wherein the respective representation of each known biological fragment is a probabilistic template, said

template providing a probability that a member of a group consisting of amino acids and nucleotides exists at a pre-determined position of said known biological fragment.

25. (Previously Presented) The method of Claim 24 wherein quantitatively determining a score of each known biological fragment in the set includes for each known biological fragment, computing the probability of existence of every subsequence of a pre-determined length in the subject genome sequence according to the probabilistic template that represents the known biological fragment.

26. (Cancelled)

27. (Currently amended) The apparatus of Claim [[26]] 37 wherein each known biological fragment in the set is represented by a respective text string.

28. (Previously Presented) The apparatus of Claim 27 wherein the scoring routine includes for each known biological fragment, counting the number of times the respective text string is found within the subject genome sequence.

29. (Currently amended) The apparatus of Claim [[26]] 37 wherein each known biological fragment in the set is represented by a probabilistic template, said template providing a probability that a member of a group consisting of amino acids and nucleotides exists at a pre-determined position of said known biological fragment.

30. (Previously Presented) The apparatus of Claim 29 wherein the scoring routine includes for each known biological fragment, computing the probability of existence of every subsequence of a pre-determined length in the subject genome sequence according to the probabilistic template that represents the known biological fragment.

31-33. (Not entered)

34. (Currently amended) A method of assigning a subject genome sequence to a class, comprising:

(a) providing a set of known biological fragments, the set being of a fixed number of said known biological fragments, each known biological fragment in the set having a respective representation;

(b) providing at least one training sequence[[s]];

(c) for each known biological fragment, quantitatively determining a score with respect to each training sequence;

(d) for each training sequence, forming a training feature vector, said training feature vector being a sequence of scores of each known biological fragment with respect to the training sequence;

(e) using the training feature vectors, classifying the training sequences, thereby defining classes of sequences;

(f) providing a subject genome sequence;

(g) quantitatively determining a score of each known biological fragment with respect to the subject genome sequence;

(h) forming a feature vector of the subject genome sequence, said feature vector being a sequence of scores of each known biological fragment in the set;

(e) using the feature vector and the training feature vectors, assigning the subject genome sequence to at least one of the defined classes of sequences, thereby producing classification, of the subject genome sequence.

35 - 36. (Cancelled)

37. (New) Apparatus for assigning a subject genome sequence to a class, comprising:

(1) an input device for inputting at least one subject genome sequence and at least one training sequence;

(2) a data store of representations of a set of a predefined number of known biological fragments; and

(3) a scoring routine executed by a digital processor having access to the data store, the scoring routine quantitatively determining a score of each known biological fragment in the set as compared against the subject genome sequence or each training sequence, said scores forming a feature vector or a training feature vector having a length equal to the predefined number of known biological sequences; and

(4) an analyzing routine executed by a digital processor, the analyzing routine performing the steps of:

(a) using the training feature vectors, classifying the training sequences, thereby defining classes of sequences; and

(b) using the feature vector and the training feature vectors, assigning the subject genome sequence to at least one of the defined classes of sequences, thereby producing classification, of the subject genome sequence.